The Impact of College on Mental Health

Lauren Phan phan.la@northeastern.edu

DS2001 - Spring 2022

Source: This project incorporates data from the Centers for Disease Control and Prevention, The National Center for Education Statistics', and the Multidiscplinary Digital Publishing Institute, all first accessed on February 28, 2022 from these respective links:

https://catalog.data.gov/dataset/mental-health-care-in-the-last-4-weeks

https://nces.ed.gov/ipeds/Search?query=fall%20enrollment%20in%20degree-granting&query2=fall%20enrollment%20in%20degree-granting&resultType=all&page=1&sortBy=relevance&overlayDigestTableId=202035

https://www.mdpi.com/2306-5729/4/3/124/htm

Abstract

In a world of quickly advancing technology and innovation, it was only a matter of time before the standards of post-secondary education raised. With these standards come high levels of stress and pressure on students; there is no denying the growing expectation for all teenagers to not just earn a degree, but to satisfy the expectations and pass the records of others. The goal of this project is to utilize raw data to further understand, investigate, and draw attention to the role that college has on student mental health. Results indicate that there is correlation between education and mental health status; regardless, this project itself cannot prove that education does negatively impact mental health for everyone.

Background

In 2017, approximately 20.4 million students in America were newly enrolled in a college or university, which is 5 million more than in 2000 (Hagan). Whether one is a first-generation citizen hoping to fulfill the American Dream or a third-generation ivy league legacy, almost all children are conditioned to believe college is the key to triumph. This immense pressure, alongside the increased toxicity of social media and the comparative nature of education, are just a few of the never-ending factors that negatively contribute to student mental health. In a study of American college students from 2020, nearly 40% claimed to have experienced depression, and one in seven had thought about suicide in the past year (Aslanian and Roth), causing negative repercussions on their headspaces, physical health, and relationships with peers and family.

Despite these drastic statistics, the mental health of students often goes unaddressed, as struggling in college to the brink of depression has unfortunately become the norm for too many

students. I am hypothesizing that post-secondary education has a significant impact on mental health, as adaptation from home, heavy coursework, and the comparative nature of education are just a few factors that can negatively impact student headspace. This project will examine the current day trends in mental health treatment alongside education and hopefully provide solid, persuasive evidence for the need for better mental health resources in schools.

More specifically, the project aims to investigate the following questions:

- Is there a trend in mental health treatments related to either age or education?
- Does college attendance have a direct relationship to the amount of mental health related illnesses in each state?
- What is the relationship between mental health status and a student's comfortability talking about mental health? How does comfortability change based on who they are confiding in?

Dataset 1

The first dataset, obtained from the CDC, comes from a Household Pulse Survey conducted by an internet questionnaire, with invitations to participate sent via email and text message. The original goal of the study was to assess the impact of Covid-19 on mental wellness. In increments of 2 weeks, it provides numbers for those who received prescription medicine, therapy, or both for mental health purposes between August 2020 and January 2022, as well as numbers for those who felt they needed therapy or medicine but did not receive it. These statistics are categorized by various characteristics; the categories most relevant to this study are the data grouped by age, education status, and state. Visualizing this data identifies trends and provides the foundation for exploring patterns in American mental health, emphasizing the treatment numbers from college students. However, when examining the following results, it is important to note and remember that these results come from an online survey; the limitations and implications of this data will be further discussed later on.

Importing Libraries and Defining Functions

In the following block of code, the necessary libraries were imported: one to read the datafile and one to create the visualizations. The datafile was also defined along with several functions that will be used throughout the project.

The first function takes a given csv file and reads it as a lists of lists, where each row is a list of its entries. This will be utilized for every dataset to make the values easy to access and work with.

In [1]:
```
# Importing Libraries and Data
import csv
import matplotlib.pyplot as plt
DATA1 = "Mental_Health_Care_in_the_Last_4_Weeks.csv"

def read_csv(filename):
    # Input: filename (a string)
    # Output: a two dimensional list of the file
```

```
    data = []
    with open(filename, "r") as infile:
        csvfile = csv.reader(infile, delimiter = ",")
        for row in csvfile:
            data.append(row)
    return data
```

Once a two dimensional list is created, a function is necessary to look through it and only utilize the ones we are interested in. Thus, the following function takes the two dimensional list, and makes a new list with data that only goes with the 3 specified variables.

In [2]:
```
def get_list_csv1(listname, category, variable1, variable2, variable3):
    # Input: listname, category, and 3 variables (all strings)
    # Output: a new list of data with only data that satisfies the variables
    lst = []
    for row in listname:
        if row[0] == variable1 and row[1] == variable2 and row[2] == variable3:
            if row[3] == category and row[9] != "":
                lst.append(float(row[9]))
    return lst
```

Given multiple lists about the same topic, the following function calculates the sum each individual list, and appends this sum into a list of all the sums.

In [3]:
```
def get_sum(individual_list, large_list):
    # Input: two lists (one that you want the sum of, and one that will contain
    # Output: list with sum of the other list appended to it
    list_sum = round(sum(individual_list), 2)
    large_list.append(list_sum)
    return large_list
```

The function below creates a bar graph given the desired values, color, labels, and title.

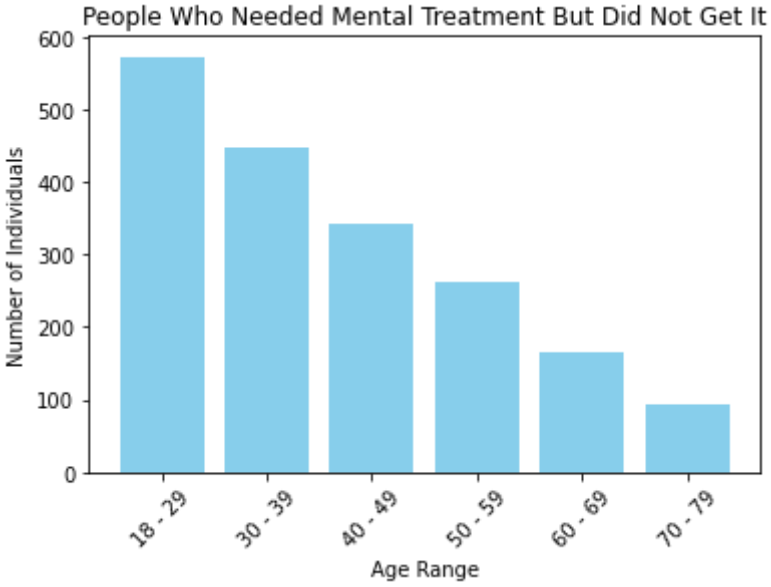
In [4]:
```
def graph(x_values, y_values, color, x_label, y_label, title):
    # Input: x and y values (lists), color (list or string), and labels and titl
    # Output: bar graph with given characteristics
    plt.bar(x_values, y_values, color = color)
    plt.xlabel(x_label)
    plt.xticks(rotation=45)
    plt.ylabel(y_label)
    plt.title(title)
    plt.show
```

Mental Health Treatment vs Age

Once the file was read and converted to a two dimensional list, it included all of the data from the survey. From this, new lists were generated by reading through the file and only including the data of interest: individuals in the United States who felt they needed therapy but did not get it, sorted by age. One list was created for each age group.

Because the data is in increments of two weeks, all of the numbers in each age list were totaled, and these age totals were added to a new list containing one value per age group. This final list was used to generate a bar graph.

In [5]:
```
# read file as 2D list
data1 = read_csv(DATA1)

# add data from each age group into its own new list
age1 = get_list_csv1(data1, "18 - 29 years", "Needed Counseling or Therapy But D
                     "By Age", "United States")
age2 = get_list_csv1(data1, "30 - 39 years", "Needed Counseling or Therapy But D
                     "By Age", "United States")
age3 = get_list_csv1(data1, "40 - 49 years", "Needed Counseling or Therapy But D
                     "By Age", "United States")
age4 = get_list_csv1(data1, "50 - 59 years", "Needed Counseling or Therapy But D
                     "By Age", "United States")
age5 = get_list_csv1(data1, "60 - 69 years", "Needed Counseling or Therapy But D
                     "By Age", "United States")
age6 = get_list_csv1(data1, "70 - 79 years", "Needed Counseling or Therapy But D
                     "By Age", "United States")

# create new list, and put the sum of each age group into this list
age_sums = []
get_sum(age1, age_sums)
get_sum(age2, age_sums)
get_sum(age3, age_sums)
get_sum(age4, age_sums)
get_sum(age5, age_sums)
get_sum(age6, age_sums)

# graph sums vs age group
x_values = ["18 - 29", "30 - 39", "40 - 49", "50 - 59", "60 - 69", "70 - 79"]
graph(x_values, age_sums, "skyblue", "Age Range", "Number of Individuals",
        "People Who Needed Mental Treatment But Did Not Get It")
```



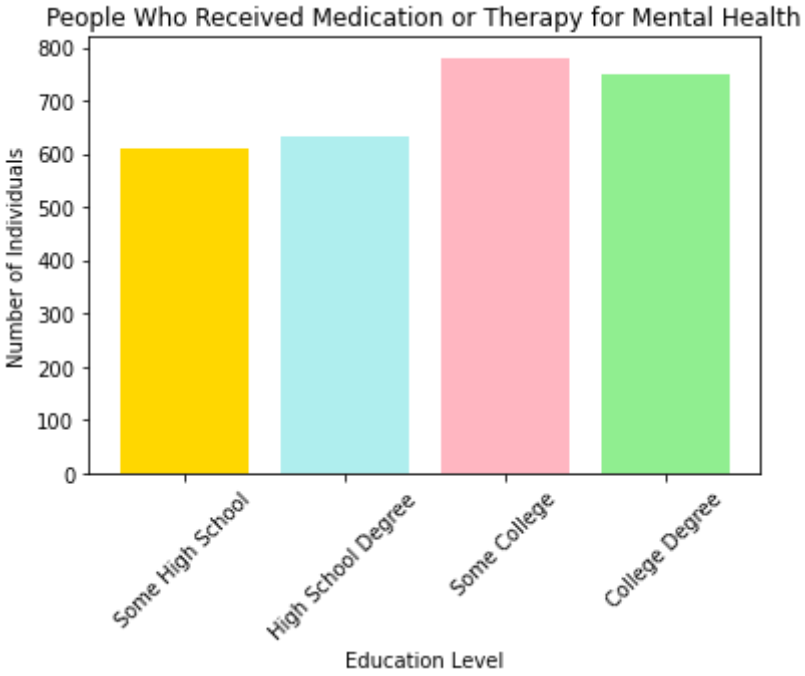
Mental Health Treatment vs Education Level

To further gauge and support education's impact on mental health, the same file will be read in a similar manner. Instead of evaluating those who needed mental health treatment but did not get it, the file will be read for those who actually took medication and/or received counseling for mental health purposes, sorted by education status.

Using the same functions and overall approach as the last graph, lists were created by reading through the file and sorting out the data by education level. The sums of these lists were then calculated and appended to a final list that was graphed.

In [6]:

```
# add data from each education level into its own list
ed1 = get_list_csv1(data1, "Less than a high school diploma",
                    "Took Prescription Medication for Mental Health And/Or Recei
                    "By Education", "United States")
ed2 = get_list_csv1(data1, "High school diploma or GED",
                    "Took Prescription Medication for Mental Health And/Or Recei
                    "By Education", "United States")
ed3 = get_list_csv1(data1, "Some college/Associate's degree",
                    "Took Prescription Medication for Mental Health And/Or Recei
                    "By Education", "United States")
ed4 = get_list_csv1(data1, "Bachelor's degree or higher",
                    "Took Prescription Medication for Mental Health And/Or Recei
                    "By Education", "United States")

# calculate the sum for each education level, and append these values to a new l
ed_sums = []
get_sum(ed1, ed_sums)
get_sum(ed2, ed_sums)
get_sum(ed3, ed_sums)
get_sum(ed4, ed_sums)

# graph sums vs education level
x_values_ed = ["Some High School", "High School Degree",
               "Some College", "College Degree"]
colors = ["gold", "paleturquoise", "lightpink", "lightgreen"]
graph(x_values_ed, ed_sums, colors, "Education Level", "Number of Individuals",
      "People Who Received Medication or Therapy for Mental Health")
```



The two graphs above show a clear trend between mental health and education. The first graph illustrates that younger adults, more of which are in college compared to older adults, have the most mental health issues that go untreated. This could be due to a variety of reasons, such as lacking time to seek mental health treatment, not having affordable, accessible treatment, or not being comfortable speaking about mental health in general, all of which align with the previous research and hypothesis of this project. The second graph also aligns with the intentions of the project, as it shows that individuals in the middle of college received the most amount of medication or therapy for mental health. This is potentially due to the large amounts of stress placed on college students, especially in comparison to high school.

Dataset 2

After identifying general trends, the dataset from the National Center for Education Statistics' collects annual data on the enrollments from every degree-granting postsecondary institute that participates in Title IV federal financial aid programs, all categorized by state. It contains data from every year between 1970 and 2019, but as the project is mainly concerned about modern day statistics, only the 2019 data is used. By combining the data from this set and the CDC dataset, college enrollment versus the number of people who received therapy or medication for mental health purposes is graphed by state to examine whether these numbers are proportional or not.

The following block of code defines a function that will create a line graph with two y axes and two lines, given 3 lists of data and 2 indeces to plot numbers between.

In [7]:
```
def line_graph(index1, index2, lst1, lst2, lst3):
    # Input: two indeces to plot between (ints) and three lists to plot, where l
    # Output: line graph with 2
    fig, ax1 = plt.subplots()
    plt.set_figwidth = 30
    ax1.set_xlabel("State")
```

```
    ax1.set_ylabel("Mental Health Treatments")
    ax1.plot(lst1[index1:index2], lst2[index1:index2], color = "blue")
    ax1.tick_params(axis = 'y', labelcolor = "blue")
    plt.xticks(rotation=45)

    ax2 = ax1.twinx()
    ax2.set_ylabel("College Enrollments")
    ax2.plot(lst1[index1:index2], lst3[index1:index2], color = "darkorange")
    ax2.tick_params(axis = "y", labelcolor = "darkorange")
    plt.title("College Enrollment and Mental Health By State")
    plt.show()
```

Once the new file is imported and read as a two dimensional list, a list of all of the states is created in alphabetical order.

In [8]:

```
# define csv file
DATA2 = "stateenrollment.csv"

# read file as 2D list
data1 = read_csv(DATA1)
data2 = read_csv(DATA2)

# create state list
all_states = ["Alabama", "Alaska", "Arizona", "Arkansas", "California", "Colorad
              "Delaware", "Florida", "Georgia", "Hawaii", "Idaho", "Illinois", "
              "Kansas", "Kentucky", "Louisiana", "Maine", "Maryland", "Massachus
              "Minnesota", "Mississippi", "Missouri", "Montana", "Nebraska", "Ne
              "New Hampshire", "New Jersey", "New Mexico", "New York", "North Ca
              "Ohio", "Oklahoma", "Oregon", "Pennsylvania", "Rhode Island", "Sou
              "South Dakota", "Tennessee", "Texas", "Utah", "Vermont", "Virginia
              "West Virginia", "Wisconsin", "Wyoming"]
```

Using the same approach as in the past two graphs, the first dataset was read to create a list of treatment numbers for each state, and the totals for each state were calcualated and added to a new list.

In [9]:

```
# make a list for each state, and put state totals into a new list
state_sums = []
for state in all_states:
    state_lst = get_list_csv1(data1, state, "Took Prescription Medication for Me
                                    " Counseling or Therapy, Last 4 Weeks", "By St
    get_sum(state_lst, state_sums)
```

Next, the alphabetical list of states will be used again to read through the second dataset of college enrollment. However, the states appear in a slightly different format than in the first. They do not have quotes, and they have an extra space at the end. Because the original state list needs to be kept for future use, a new list was created to edit these names as they appear in the second dataset.

Once the list is created, the enrollment dataset will be read through to gather data of student enrollment in each state.

In [10]:

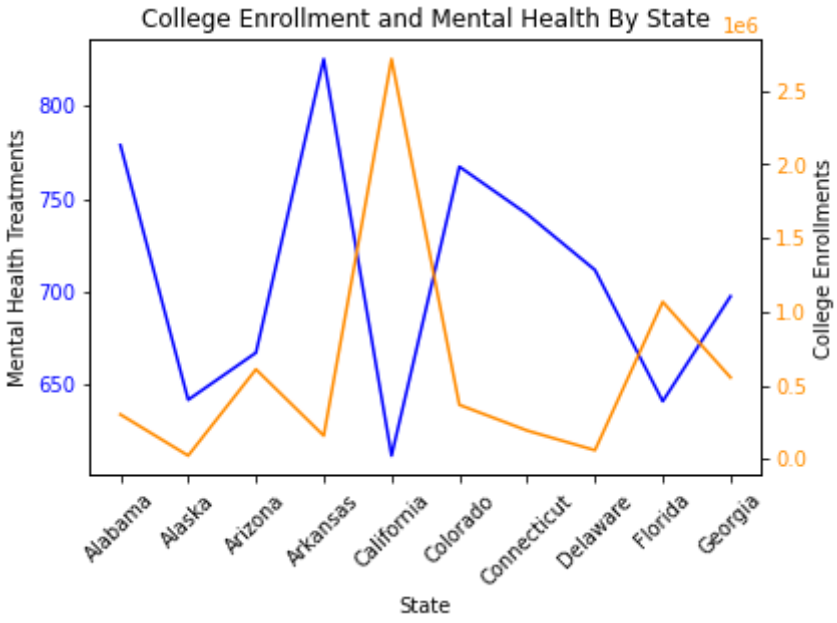
```
# make a new list to remove quotes and add a space to state names
```

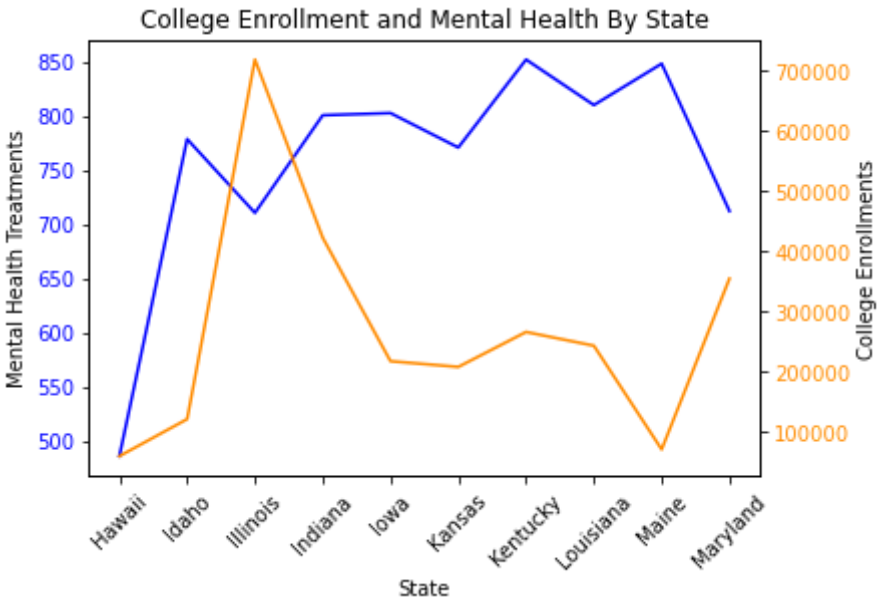
```
clean_names = []
for name in all_states:
    quoteless_name = name.replace('"', "")
    clean_name = quoteless_name + " "
    clean_names.append(clean_name)

# create a list with enrollment data for each state only from 2019
enrollment_lst = []
for row in data2:
    if row[0] in clean_names:
        enrollment_lst.append(row[10])
```

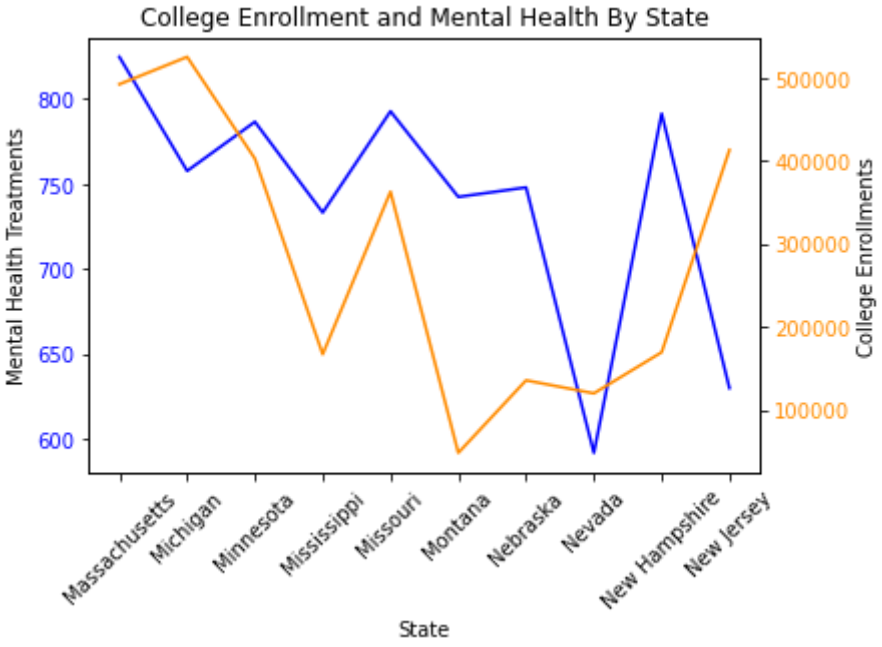
To graph these numbers and have the program read them properly, they must first be converted from strings to integers. The following block of code does this conversion, and then graphs the results. The graph has been divided into smaller subgraphs for easier analyzing.

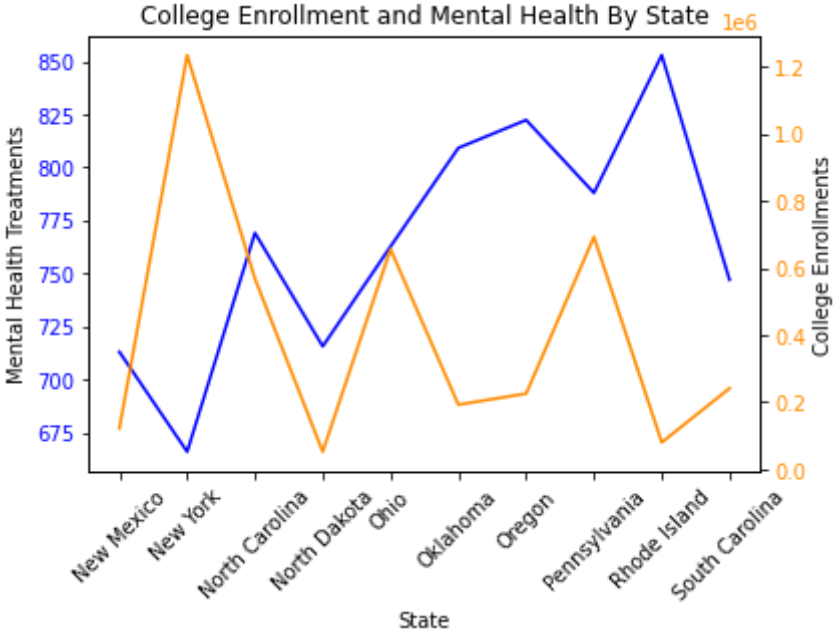
In [11]:
```
# convert enrollment data from string to integer
enrollment = []
for num in enrollment_lst:
    new_num = num.replace(",", "")
    enrollment.append(int(new_num))
```

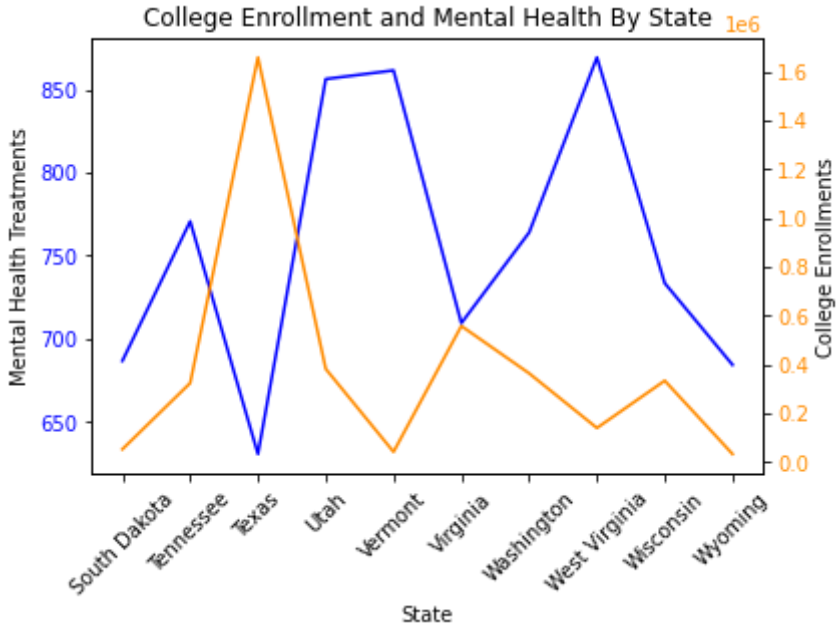
In [12]:
```
# graph enrollment data vs mental treatment by state
line_graph(0, 10, all_states, state_sums, enrollment)
line_graph(10, 20, all_states, state_sums, enrollment)
line_graph(20, 30, all_states, state_sums, enrollment)
line_graph(30, 40, all_states, state_sums, enrollment)
line_graph(40, 50, all_states, state_sums, enrollment)
```




College Enrollment and Mental Health By State
Mental Health Treatments
College Enrollments
850
800
750
700
650
600
550
500
700000
600000
500000
400000
300000
200000
100000
Hawaii
Idaho
Illinois
Indiana
Iowa
Kansas
Kentucky
Louisiana
Maine
Maryland
State


College Enrollment and Mental Health By State
Mental Health Treatments
College Enrollments
800
750
700
650
600
500000
400000
300000
200000
100000
Massachusetts
Michigan
Minnesota
Mississippi
Missouri
Montana
Nebraska
Nevada
New Hampshire
New Jersey
State


College Enrollment and Mental Health By State
1e6
Mental Health Treatments
College Enrollments
850
825
800
775
750
725
700
675
1.2
1.0
0.8
0.6
0.4
0.2
0.0
New Mexico
New York
North Carolina
North Dakota
Ohio
Oklahoma
Oregon
Pennsylvania
Rhode Island
South Carolina
State



Although it was expected that mental health treatments in young adults and college enrollments would have a directly proportional relationship, the graphs created do not truly support or deny this. While some states do, others do not, and it is not clear whether it is due to purely chance or not. Moreover, while these graphs signify that there is no direct relationship among those that participated in the survey, they cannot be used to draw conclusive relationships about the overall population, as the dataset gave no information on its coverage. In other words, it is unknown how many people they surveyed as a whole, as well as whether they surveyed an even proportion of people per state. If every state was not covered equally, that could have altered the data.

Dataset 3

The dataset from the Multidisciplinary Digital Publishing institute comes from an online questionnaire given to students at an international university in Japan. The original goal of the survey was to compare the mental healths of international versus domestic students, as they predicted that international students face more mental health issues during the transition to college. The set includes data on each student's stress levels, symptoms of depression, and likelihood of reaching out to peers and family to talk about mental health. Comparing students' likelihood to reach out to peers versus family members, professors, counselors, and doctors can identify possible ways to make mental health a more important conversation topic.

First, the functions used for this dataset are defined. The following function takes a two dimensional file and appends the data from a given category; the function after that does the same thing, but converts the data into integers, too.

In [13]:
```
def get_dep_lst(filename, category):
    # Input: a datafile (list) and a specific category (string)
    # Output: a new list with data only from given category
    lst = []
    for row in filename:
        if row[18] == category:
```

```
            lst.append(row)
    return lst
```

In [14]:

```
def get_comfort_lst(dep_lst, relationship_index):
    # Input: a list and an integer
    # Output: a new list, with data from the given list only at the given index
    lst = []
    for row in dep_lst:
        value = int(row[relationship_index])
        lst.append(value)
    return lst
```

The final function calculates the average of a given list and adds it to another list.

In [15]:

```
def get_avg(individual_list, large_list):
    # Input: two lists
    # Output: the average of one list appended to the other list
    list_avg = sum(individual_list) / len(individual_list)
    large_list.append(list_avg)
    return large_list
```

First, the datafile is imported as well as NumPy, which will be used to arrange numbers later on. Once the file is read, separate lists are created to sort the file based on the individual's depression level. This depression level was measured via the Patient Health Questionnaire PHQ-9, which is a commonly used screening tool that examines depression symptoms and severity. This is done by listing nine symptoms and having the participants rank how often they have experienced the symptom throughout the past 2 weeks. The depression severities that will be examined are mild, moderate, and severe.

In [16]:

```
DATA3 = "data.csv"
import numpy as np

# read file as a 2D list
data3 = read_csv(DATA3)

# create separate lists based on depression level
mild_depress = get_dep_lst(data3, "Mild")
mod_depress = get_dep_lst(data3, "Mod")
sev_depress = get_dep_lst(data3, "Sev")
```

In one component of the survey, patients were asked how likely they were to seek help from different specific people in the event of a mental health emergency on a scale from 0 to 7, such as parents, friends, professors, phone hotlines, and doctors or therapists. To gauge who students are comfortable speaking towards mental health about, these comfortability values will be compared among the different depression levels.

The following block of code creates lists for the comfortability of students speaking with each person and calculates the average comfortability, all done separately for the three depression levels.

In [17]:

```
# for each depression level, create lists of comfortability speaking with each p
# the values, and append that into a list
mild_parents = get_comfort_lst(mild_depress, 30)
mild_friends = get_comfort_lst(mild_depress, 29)
mild_profess = get_comfort_lst(mild_depress, 32)
mild_phone = get_comfort_lst(mild_depress, 33)
mild_doctor = get_comfort_lst(mild_depress, 34)
mild_alone = get_comfort_lst(mild_depress, 36)
mild_lst = []
milds = [mild_parents, mild_friends, mild_profess, mild_phone, mild_doctor, mild
for lsts in milds:
    get_avg(lsts, mild_lst)

mod_parents = get_comfort_lst(mod_depress, 30)
mod_friends = get_comfort_lst(mod_depress, 29)
mod_profess = get_comfort_lst(mod_depress, 32)
mod_phone = get_comfort_lst(mod_depress, 33)
mod_doctor = get_comfort_lst(mod_depress, 34)
mod_alone = get_comfort_lst(mod_depress, 36)
mod_lst = []
mods = [mod_parents, mod_friends, mod_profess, mod_phone, mod_doctor, mod_alone]
for lsts in mods:
    get_avg(lsts, mod_lst)

sev_parents = get_comfort_lst(sev_depress, 30)
sev_friends = get_comfort_lst(sev_depress, 29)
sev_profess = get_comfort_lst(sev_depress, 32)
sev_phone = get_comfort_lst(sev_depress, 33)
sev_doctor = get_comfort_lst(sev_depress, 34)
sev_alone = get_comfort_lst(sev_depress, 36)
sev_lst = []
sevs = [sev_parents, sev_friends, sev_profess, sev_phone, sev_doctor, sev_alone]
for lsts in sevs:
    get_avg(lsts, sev_lst)
```

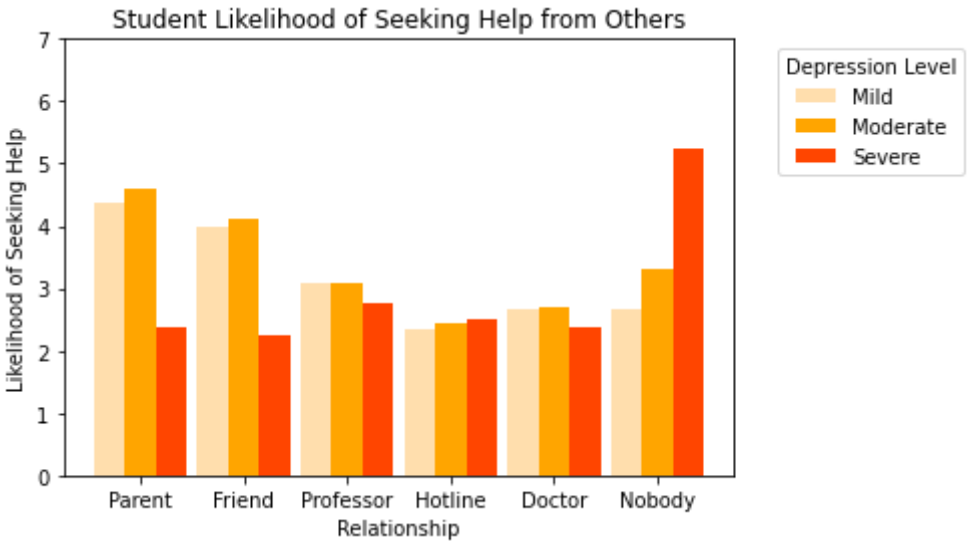
Once these values are separated by depression type and relationship, and the averages are calculated, the following block of code is used to generate a bar graph.

In [18]:

```
# graph the data
relationships = ["Parent", "Friend", "Professor", "Hotline", "Doctor", "Nobody"]
X_axis = np.arange(len(relationships))
plt.bar(X_axis - 0.3, mild_lst, 0.3, label = 'Mild', color = 'navajowhite')
plt.bar(X_axis , mod_lst, 0.3, label = 'Moderate', color = 'orange')
plt.bar(X_axis + 0.3, sev_lst, 0.3, label = 'Severe', color = 'orangered')
plt.ylim(0, 7)
plt.xlabel("Relationship")
plt.ylabel("Likelihood of Seeking Help")
plt.title("Student Likelihood of Seeking Help from Others")
plt.xticks(X_axis, relationships)
plt.legend(title = "Depression Level", bbox_to_anchor=(1.05, 1))
```

Out[18]:

```
<matplotlib.legend.Legend at 0x7fec4011a6a0>
```



The following visualization shows that for students in this survey, comfortability seeking mental health aid is significantly lower for those with severe depression compared to mild and moderate. This applies to every relationship except an emergency hotline, where those with higher levels of depression are slightly more likely to seek help. Another interesting trend is that for parents and friends, those with severe depression are much less likely to reach out, which is not generally expected.

The most notable trend is that those with higher levels of depression are more likely to not seek help at all; this draws attention to mental health being a taboo for many people, especially students who feel pressure to succeed. It also supports the need for educational institutions to establish more accessible, clear mental health resources for their students, as well as people being aware of student mental health, its symptoms, and resources as a whole.

Conclusions and Ethical Implications

Many of the graphs did support the notion of education causing negative effects on mental health. The first dataset illustrated that mental health cases were the highest for young adults and those in college. The second dataset illustrated no direct correlation between college and mental health, but because of the limitations of the survey, the possibility of a direct correlation cannot be confirmed or denied. The last dataset showed that in general, comfortability seeking mental help aid decreases for those with more severe depression.

Most of the data in this project came from surveys; one was conducted online via phone and text, while another was conducted at a university. The initial purpose of examining survey data was rooted in the notion that mental health is often viewed as a taboo in society, which is supported by the final bar graph. Because the surveys are anonymous, the hypothesis was that the answers were probably more honest than they would have been if the particpants had to provide their identity alongside their answers, although this cannot be confirmed or denied.

However, it is important to note that because they are surveys, there are limitations unrelated to mental health that could also be influencing the results. For the online survey especially, there is no information regarding how the survey was sent out or what population it reached. For

example, in consideration of the graph that examined age groups, it is possible that there were simply more people in college that responded, and the data may be skewed as a result of that. Thus, while there are certainly correlations present that can aid our hypothesis, these correlations do not indicate causation; the results of the data are limited only to those that responded and cannot be assumed to fit the entire population.

Nevertheless, these graphs do fulfill the original intent of drawing attention to the lack of and need for more mental health awareness among students, families and friends of students, and education faculty and administrators. Mental health cases are higher now than ever, and it is predicted that without more awareness and education about the topic, these numbers will not decrease. To aid this rising issue, educational institutions should educate more on the symptoms of and treatment for common mental health issues, not just among their students, but also to the students' families. Therapy and relaxing recreational activities should be accessible, affordable, and heavily communicated at all post secondary instutions. Most importantly, it is imperative that students not only know where to go in case of a mental health emergency, but also that they feel comfortable reaching out and speaking about the topic. Thus, families and instutitions should prioritize fostering an open, communicative environment where school-community collaboration is a core pillar. Ultimately, while education is certainly at the forefront of evolving society, it should never be achieved at the expense of anyone's health and happiness.

References

Aslanian, Sasha, and Alisa Roth. “Inside the College Mental Health Crisis.” Inside the College Mental Health Crisis | Under Pressure | APM Reports, APM Reports, 6 Jan. 2022, https://www.apmreports.org/episode/2021/08/19/under-pressure-the-college-mental-health-crisis.

Hagan, Ekua. “Is College Necessary?” Psychology Today, Sussex Publishers, 2 Apr. 2018, https://www.psychologytoday.com/us/blog/teen-angst/201804/is-college-necessary.

“Mental Health Care in the Last 4 Weeks.” Mental Health Care in the Last 4 Weeks - CKAN, Publisher Centers for Disease Control and Prevention, 11 Jan. 2021, https://catalog.data.gov/dataset/mental-health-care-in-the-last-4-weeks.

Nguyen, Minh-Hoang, et al. “A Dataset of Students' Mental Health and Help-Seeking Behaviors in a Multicultural Environment.” MDPI, Multidisciplinary Digital Publishing Institute, 21 Aug. 2019, https://www.mdpi.com/2306-5729/4/3/124/htm.

“Postsecondary School Locations 2019-20.” Postsecondary School Locations 2019-20 - CKAN, Publisher NCES, 25 Jan. 2022, https://catalog.data.gov/dataset/postsecondary-school-locations-2019-20.

In [ ]: